DOCUMENT RESUME

ED 148 713 SO 010 646

TITLE Evaluation and Social Studies: Practical Ideas for Classroom Teachers.
INSTITUTION Texas Education Agency, Austin. Div. of Curriculum Development.
PUB DATE [76]
NOTE 49p.; For a related document, see SO 010 647

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Academic Achievement; Affective Objectives; Cognitive Objectives; Criterion Referenced Tests; Educational Assessment; Elementary Secondary Education; *Evaluation; *Evaluation Methods; *Feedback; *Guidelines; Measurement; Multiple Choice Tests; Norm Referenced Tests; Program Evaluation; Psychomotor Objectives; Skill Development; *Social Studies; *Test Construction; Testing; Test Items

ABSTRACT

The bulletin is designed to help educators understand the rationale and methods for instituting continuous evaluation of social studies teaching and learning in terms of specified objectives. Chapter one explains how evaluation in social studies is helpful to teachers, students, school officials, and parents. Evaluation includes testing as well as classroom observations, group discussions, diaries, sociograms, and interviews. An effective program must be based on clear objectives, be continuous, and be interpreted correctly to all groups involved. Chapter two explains how to evaluate the instructional process. It discusses how to clarify cognitive, affective, and psychomotor objectives, identify specific skills to be evaluated, and determine the most appropriate type of tests. Guidelines for writing short-answer items, multiple-choice items, true-false items, matching items, and essay questions are given in chapter three. Chapter four describes how to provide useful feedback after evaluation components have been completed. Two appendices list 16 references about evaluation and 26 important terms. (AV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED148713

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Texas Education Agency
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATION AND SOCIAL STUDIES: PRACTICAL IDEAS FOR CLASSROOM TEACHERS

Texas Education Agency
Division of Curriculum Development

SP 010 646

After the original free distribution, additional copies may be purchased for $1.00; however, with the increasing demand for educational material, the supply at times may be exhausted.

Since Texas Education Agency publications are not copyrighted, any or all sections of this bulletin may be duplicated.

Agency bulletins are distributed to the offices of school district and county superintendents; these publications are the property of those offices and are not to be considered as personal property of the individual who receives them.

COMPLIANCE WITH TITLE VI, CIVIL RIGHTS ACT OF 1964; TITLE IX OF THE EDUCATION AMENDMENTS OF 1972; AND THE MODIFIED COURT ORDER, CIVIL ACTION 5281, FEDERAL DISTRICT COURT, EASTERN DISTRICT OF TEXAS, TYLER DIVISION

Reviews of local education agencies pertaining to compliance with Title VI Civil Rights Act of 1964; Title IX of the Education Amendments of 1972; and with specific requirements of the Modified Court Order, Civil Action No. 5281, Federal District Court, Eastern District of Texas, Tyler Division are conducted periodically by staff representatives of the Texas Education Agency. These reviews cover at least the following policies and practices:

(1) acceptance policies on student transfers from other school districts;

(2) operation of school bus routes or runs on a non-segregated basis;

(3) non-discrimination in extracurricular activities and the use of school facilities;

(4) non-discriminatory practices in the hiring, assigning, promoting, paying, demoting, reassigning, or dismissing of faculty and staff members who work with children;

(5) enrollment and assignment of students without discrimination on the basis of race, color, national origin, or sex;

(6) non-discriminatory practices relating to the use of a student's first language; and

(7) evidence of published procedures for hearing complaints and grievances.

In addition to conducting reviews, the Texas Education Agency staff representatives check complaints of discrimination made by a citizen or citizens residing in a school district where it is alleged discriminatory practices have occurred or are occurring.

Where a violation of Title VI of the Civil Rights Act and/or Title IX of the Education Amendments of 1972 is found, the findings are reported to the Office for Civil Rights, Department of Health, Education and Welfare.

If there is a direct violation of the Court Order in Civil Action No. 5281 that cannot be cleared through negotiation, the sanctions required by the Court Order are applied.

FOREWORD

EVALUATION AND THE SOCIAL STUDIES: PRACTICAL IDEAS FOR CLASSROOM TEACHERS was developed for use primarily by the classroom teacher. However, the evaluation of the teaching-learning process affects everyone in the educational environment, and the bulletin will be of interest to administrators and parents alike. The bulletin is being distributed to school superintendents with sufficient copies for elementary and secondary school campuses, as well as to Education Service Centers and heads of university departments of education.

The bulletin may be used on an individual basis or in staff meetings, as information in educational courses or for discussion, and as a springboard for reviewing how social studies instruction can be evaluated.

It is hoped that this bulletin will provide assistance to all those interested or concerned with the evaluation of teaching and learning in the social studies.

M. L. Brockette
Commissioner of Education

TEXAS EDUCATION AGENCY

STATE BOARD OF EDUCATION
(State Board for Vocational Education)

MRS. ELIZABETH R. ARMSTRONG, Lake Jackson

VERNON BAIRD, Fort Worth

WOODROW W. BEAN, El Paso

JAMES M. BINION, Abilene

JOE KELLY BUTLER, Houston

W. W. CARTER, Amarillo

DON H. COOK, D.D.S., Mesquite

BRYAN CROUCH, Poteet

DR. VIRGINIA CURREY, Arlington

JIMMY L. ELROD, San Antonio

W. H. FETTER, La Marque

WAYNE FREDERICK, Orange

JOSEPH C. GATHE, M.D., Houston

E. R. GREGG, JR., Jacksonville

RUBEN E. HINOJOSA, Mercedes

JIM HOLLINGSWORTH, De Soto

WILLIAM N. KEMP, O.D., Houston

MRS. MARY ANN LEVERIDGE, East Bernard

PAUL MATHEWS, Greenville

DR. STANLEY B. McCALEB, Richardson

GLEN L. SMITH, Waco

MRS. YOLANDA P. TORRES, San Antonio

MRS. JANE H. WELLS, Austin

JAMES H. WHITESIDE, Lubbock

M. L. BROCKETTE, Commissioner of Education
(Executive Officer of the State Board of Education)

L. HARLAN FORD, Deputy Commissioner
for Programs and Personnel Development

ALTON O. BOWEN, Deputy Commissioner
for Administrative Services

DOROTHY DAVIDSON, Associate Commissioner
for General Education

LEROY F. PSENCIK, Director
Division of Curriculum Development

SOCIAL STUDIES STAFF

LOUIS GRIGAR, Program Director

Consultants

LARRY PERRY

ROBERTA ANN RAITHEL

ELVIN TYRONE

ACKNOWLEDGEMENTS

Appreciation is extended to the following individuals for their cooperation in the development of this bulletin.

Linda Avena	Parent San Antonio Independent School District
Diane Greer	Teacher Lamar Junior High School Austin Independent School District
William Guardia	Social Studies Consultant San Antonio Independent School District
Joe Hegarty	Consultant for Evaluation Texas Education Agency
Irene Kanter	Teacher Anderson High School Austin Independent School District
James Kracht	Associate Professor of Curriculum and Instruction Texas A&M University
Lola LeBlanc	Social Studies Consultant Corpus Christi Independent School District
Doris Owens	Teacher Harris Elementary School Austin Independent School District
Marvin Veselka	Director of Program Evaluation Texas Education Agency

CONTENTS

INTRODUCTION 1
PURPOSES OF EVALUATION IN SOCIAL STUDIES 2
EVALUATING THE INSTRUCTIONAL PROCESS 7
IMPROVEMENT OF SELECTED EVALUATION TOOLS 20
USING THE DATA GAINED FROM EVALUATION TO IMPROVE INSTRUCTION 34
APPENDIX A
General References 38
Selected References 38
APPENDIX B
Glossary of Terms 40

INTRODUCTION

One of the most pressing problems facing the social studies educator is the evaluation of the learning process. To some social studies teachers, the way to evaluate student progress is to give a series of exercises and tests, mark them, add or average the scores, and enter them on a report card. The results of this process, although simple and fast, have not always proved satisfactory in evaluating either the student or the instructional program.

The need to develop realistic and effective programs of evaluation received increased impetus when the State Board of Education in 1970 included in its *Goals for Public School Education in Texas* a section promoting accountability in the public schools. Each local school district was to establish a program of continuous planning and evaluation "for measuring the performance of the public school system in terms of the competence of its staff, the performance of its pupils, and the efficiency of its structure and process." The Texas Education Code provides that programs and services in schools should be suitable and appropriate to the educational needs of the students. These actions point up the fact that all educators, including those in social studies, must be accountable for the effectiveness of local educational programs.

The Texas Education Agency bulletin, *Framework for the Social Studies, Grades K-12*, states that evaluation of teaching and learning is a continuous process of diagnosing, appraising, and measuring the effectiveness of the program in terms of the aims of the curriculum and the specific objectives for a grade-level study, a course, or a unit.

Evaluation and the Social Studies: Practical Ideas for Classroom Teachers is intended as a support bulletin for the *Framework*. This bulletin is designed to capsulize much of the growing body of research for the busy educator who is too hard-pressed for time to collect and review the literature relative to evaluation and the social studies.

PURPOSES OF EVALUATION IN SOCIAL STUDIES

Evaluation of the learner is an essential step in the teaching and learning process. Through a comprehensive evaluation program both the student and the teacher are involved in educational activities that help ensure optimal learning. This process provides teachers information on the effectiveness of the instructional program and indicates to students their strengths and weaknesses in the acquisition of knowledge and skills.

The assessment of educational outcomes is essential for sound planning in curriculum development. The responsibility to evaluate whether educators and students are achieving educational objectives is one shared by all individuals involved in the teaching-learning process:

Students. At nearly every point during their education, students want to know, and have a right to know, how they are doing in their schooling. Questions that students might ask are:

- What are my strengths and weaknesses?
- Am I meeting my expectations for myself?
- Am I meeting the teacher's expectations?
- Am I developing feelings of satisfaction and accomplishment?
- Am I becoming qualified for the next important step in my schooling?
- How am I doing in comparison to my class or group?

Teachers. A teacher may have several needs for the evaluation of the learner. One comes from concern about the success of teaching while a second comes from concern with a specific student's success or failure. Teachers may also want to compare the results of local instructional programs with results from similar programs in other parts of the state or nation. Questions that teachers might ask are:

- Are the objectives of the course of study appropriate?

- Have the students met the objectives?
- What appear to be the points that demand additional instruction?
- What methods can best bring about improvements in learning?
- How can additional or remedial instruction be provided to meet a specific student's needs?

School officials. A comprehensive evaluation of the social studies program should reveal not only the strengths but also areas of needed improvements. Questions that school officials might ask are:

- What are the strengths and weaknesses of the social studies program?
- Is the current social studies program achieving district wide goals and objectives?
- What type changes and/or experimentations should be conducted to improve the instructional program?
- What are the educational gains of our students as compared to students in other districts?
- What are the strengths and weaknesses of the social studies teachers?
- What inservice programs appear to be needed?

Parents. As long as school is seen as a means of advancement or a way of securing a place in society, and as long as parents have hopes and fears for their children, they will be concerned about their educational progress. A school evaluation program should therefore involve parents in the process as educational partners. Questions that parents might ask are:

- How good is the local social studies program?
- Are students gaining in skills and knowledge in the program?
- Are different educational needs of all students being met by the program?

- Are teachers trying different methods of teaching to meet the needs of the students?

- Are the evaluation data used to help the students progress in the instructional program?

- What are ways that the evaluation data can be used by the parent to help the teacher help the student?

Many individuals and groups have legitimate interests in how students are progressing and in the kind of job the schools are doing in meeting the educational needs of the students. It is important for the social studies educator to remember that the purpose of evaluation is to provide feedback and guidance to the whole educational process at every level.

EVALUATION AS AN INSTRUCTIONAL TOOL

There are many facets to evaluating growth in learning. Many teachers, parents, and administrators tend to think of the measurement of educational growth in terms of tests and testing only. *Evaluation* is a collective term for many of the ways and methods available for students, teachers, administrators, and parents to obtain feedback regarding the effectiveness of the teaching-learning process. *Testing* is only a component of evaluation, a component which involves the use of a particular grouping of instruments and/or procedures to determine students' knowledge or intelligence. *Measurement* is the procedure by which student behavior is described, usually in mathematical terms. *Grading* is the attempt by teachers to sum up their estimates of students and their work in terms of marks or symbols. Thus, evaluation becomes the umbrella under which these components operate in the instructional program.

Evaluation must be diversified and an integral part of all stages of curriculum planning and implementation. Within the classroom, student evaluation by teachers (and specialists) may include:

- classroom observations

- interviews and critiques

- tape recordings of individual and group activities

- case studies
- students' check lists and work sheets
- group discussions
- individual and group conferences
- diaries and logs
- sociograms
- the use of pictures, drawings, murals, and montages
- checklists, inventories, and questionnaires
- audio and visual tapes
- commercial tests
- teacher-made tests

CHARACTERISTICS OF AN EFFECTIVE EVALUATION PROGRAM IN THE SOCIAL STUDIES

A major function of evaluation as an instructional tool is to demonstrate to the social studies educator the effectiveness of the instructional program. An effective program of evaluation is many-faceted. Characteristics are that the program:

- is based on clear and specific objectives
- is continuous and cumulative, with records kept by teachers and students
- includes as much subjective material as possible, provided by teachers and others
- includes as many objective methods of evaluation as possible
- is carried on daily through informal methods and also at frequent intervals with more formal instruments

- includes planning by teachers so that evaluation data can be utilized fully and effectively
- is wisely interpreted to students, parents, and others
- does not unduly stress any one socioeconomic class
- includes controlled teacher experimentation with new methods and devices
- includes or involves students, parents, and administrators

For evaluation to serve as an instructional tool, there must be some appraisal of the extent to which the program objectives and student progress are achieved. This process involves student and teacher self-evaluations and not only provides a blueprint of strengths and weaknesses, both for the individual and the group, but also serves as a guide for future planning by students, teachers, administrators, and parents.

EVALUATING THE INSTRUCTIONAL PROCESS

Student evaluation is a pointless process unless it is directed toward assessing progress toward goals defined by objectives. All too frequently goals and objectives in social studies, as in other content areas, are stated by educators and then forgotten. Teachers and administrators know that statements of goals, no matter how lucid and appropriate, remain statements on paper until they are made operational in the lessons children learn and through the opportunities provided students for practicing what they have learned.

Evaluation is useful in the formulation and modification of goals, objectives, and instructional procedures. Evaluation also functions as a feedback mechanism to the educational community. This process is helped by the use of instructional objectives.

USE OF INSTRUCTIONAL OBJECTIVES

Instructional objectives that identify the specific student behavior that is sought are of special value to the teacher in selecting and evaluating learning experiences. If the behavior of the student can be observed while demonstrating the acquired knowledge, or if the product of the student knowledge is observable, then the objective is acceptable.

An instructional objective should consist of three parts:

- What the student will do (construct, write, cite, etc.)
- Definition of important conditions (restrictions and limitations)
- Criterion of acceptable performance (five out of 10, etc.)

The following instructional objectives illustrate the characteristics that make them so important to the process of evaluation:

- Given a list of 20 foods, the student will identify five kinds of food eaten by the early Tejas Indian.
- The student will be able to name three distinct kinds of climates found in Texas and will point out their general geographical areas on a large map of Texas.

- Given a list of factors leading to significant historical events, the student must be able to select at least five factors contributing to the depression of 1929.

- The student after completing a reading selection on the American Indian, will describe orally or in writing, problems encountered in our society by minority groups such as the American Indian.

Social studies classroom teachers generally work with two kinds of instructional objectives:

- Those applicable to a given unit of work. (The combined number of units taught during the school year provide a list of yearly objectives.)

- Those written for each individual lesson. (These would designate detailed and specific daily objectives to be achieved by the learner.)

If teachers can take related objectives, identify, and group them under appropriate headings, they become more manageable, useful, and valuable as means of improving content, instruction, and evaluation.

A taxonomy of educational objectives, developed by Benjamin Bloom and his associates, covering both the cognitive and affective domains, offers one classification system of objectives. This particular taxonomy classifies educational objectives in terms of the behavior of the learners. The importance of the taxonomy for social studies educators is that it illustrates the hierarchical nature of intellectual behavior, helps identify the specific learner outcomes, and underscores the fact that neither instruction nor evaluation should deal exclusively with factual details. The classes of behavior are arranged in hierarchical order and include broad categories of behavior: (1) cognitive (dealing with knowledge and skills), (2) affective (dealing with values and attitudes), and (3) psychomotor (dealing with motor acts and developing skills). On pp. 9-10 is a model of Bloom's taxonomy of cognitive objectives, definitions, and examples of questions illustrating the classes of behavior students should be able to exhibit.

TAXONOMY OF EDUCATIONAL OBJECTIVES
(Cognitive Domain)

Class of Behavior	Definition	Examples
1. Knowledge	Recalls or recognizes information. Has knowledge of major ideas.	WHAT IS THE NAME OF THE CAPITAL OF THE UNITED STATES?
2. Translation	Changes information into a different symbolic form or language. Interprets communications into major ideas and their interrelationships.	YOU HAVE READ AND DISCUSSED THE DECLARATION OF INDEPENDENCE AND ITS IMPORTANCE TO EVENTS TODAY. WRITE THE SECOND SENTENCE OF THE DECLARATION OF INDEPENDENCE IN YOUR OWN WORDS.
3. Interpretation	Discovers relationships among facts, generalizations, definitions, values, and skills. Finds the relationship among the parts of the subject matter being studied.	WE HAVE BEEN STUDYING THE TERM, *MONOPOLY*. WRITE A PARAGRAPH ON EACH OF THE FOLLOWING TELLING THE EXTENT TO WHICH EACH IS A MONOPOLY IN OUR COMMUNITY. A. GARBAGE COLLECTION B. BARBERING SERVICE C. MOVING PICTURE ENTERTAINMENT D. HOSPITAL CARE
4. Application	Solves a lifelike problem that requires the identification of the issue and the selection and use of appropriate generalizations and skills. Transfers previous knowledge to new situations in which appropriate facts, generalizations, values, and skills are used to solve the new problem.	WE HAVE STUDIED THE MAIN REGIONS OF THE UNITED STATES. THIS INCLUDED CLIMATE, TOPOGRAPHY, CROPS, POPULATION, VEGETATION, MANUFACTURING, AND AGRICULTURE. HERE ARE A DOZEN BIG LANDSCAPE PICTURES. STUDY EACH PICTURE CAREFULLY AND THEN NAME A STATE IN WHICH YOU THINK THE PICTURE MIGHT HAVE BEEN TAKEN. GIVE AS MANY REASONS AS YOU CAN WHY YOU THINK YOUR LOCATION MIGHT BE IN THE STATE YOU NAME.

TAXONOMY OF EDUCATIONAL OBJECTIVES, cont'd.

Class of Behavior	Definition	Examples
5. Analysis	Solves a problem in the light of conscious knowledge of the parts and forms of thinking. Detects fallacies, supportive statements, relevant and irrelevant statements to an argument.	LISTEN TO THIS STORY AND DECIDE WHAT IS WRONG WITH JOHN'S CONCLUSION. "ON THREE SUCCESSIVE FISHING TRIPS, JOHN WORE HIS OLD BATTERED FELT HAT AND HAD AMAZINGLY GOOD LUCK. ON THE FOURTH TRIP HE DIDN'T WEAR IT AND RETURNED EMPTY-HANDED. NOW HE WON'T THINK OF GOING FISHING WITHOUT IT."
6. Synthesis	Solves a problem that requires original, creative thinking. Has ability to create, describe, theorize, discover, or devise tests of hypotheses in creating an original product.	WE HAVE BEEN STUDYING THE EVENTS THAT SURROUNDED THE BOSTON MASSACRE. WRITE A PARAGRAPH TELLING WHAT EACH OF THESE PEOPLE MIGHT HAVE SAID WHEN THEY HEARD THE NEWS OF THE BOSTON MASSACRE. A. A TORY B. A MEMBER OF THE SONS OF LIBERTY C. A BRITISH SOLDIER
7. Evaluation	Makes judgments toward some end about the value of ideas, methods, or materials. Has ability to apply internal, external, comparative, and aesthetic standards as criteria in solving problems.	WE HAVE BEEN STUDYING ABOUT THE INCREASING NUMBER OF WOMEN WHO ARE ENTERING THE JOB MARKET. IF THE HUSBAND OF A FAMILY EARNS A LIVING WAGE, IS IT FAIR FOR THE WIFE TO SEEK EMPLOYMENT? SUPPORT YOUR ANSWER.

IDENTIFYING THE SKILLS TO BE EVALUATED

Social studies educators must be aware of the role of social studies skills as they work with students day-to-day and as they help prepare students to be active participants in evaluating their own educational progress. Social studies skills are an essential part of any social studies program and are basic to the teaching-learning process. They include:

- **Study and Research**—these skills are invaluable if students are to make efficient and wise use of resources in fact-gathering activities. These activities are necessitated by students' desire to answer questions and to solve problems. Examples of study and research skills include the ability to:
  - Locate and gather information
  - Do critical reading, writing, discussing, and listening
  - Select relevant and accurate information
  - Organize information around a significant problem
- **Problem-Solving**—these skills are essential in the development of units of study. Examples of these skills include the ability to:
  - Recognize a problem area
  - Identify specific problems within the area
  - Formulate a hypothesis
  - Gather data to support, refine, or reject the hypothesis
  - Draw conclusions
  - Apply conclusions to current and future problems
- **Critical Thinking**—these skills are extremely important in the problem-solving process. Examples are being able to:
  - Evaluate sources of data
  - Interpret and synthesize data from a variety of sources and experiences
  - Base conclusions on data
  - Change conclusions as new data are found

- **Group-Process**—these skills are an essential component for the learner. Examples are having the ability to do:

  - Cooperative planning
  - Group work
  - Reporting and discussion of data
  - Evaluation (group and self)

TESTING IN THE EVALUATION PROCESS

Testing is a component of evaluation which involves the use of a particular grouping of instruments and/or procedures. An evaluation program requires well conceived and developed tests. As most social studies educators know, preparing such tests is not a simple task. It means that social studies educators must have some detailed knowledge of the mechanics of test designs, comprehension of the various types of tests, understanding of the several types of questions by format, and competence in writing both cognitive and affective questions.

One of the key concepts that social studies educators must remember is that students should be offered a wide variety of test situations. There should not be complete dependence on the paper-and-pencil test as a means of evaluating the objectives of the course. Testing situations may include:

- the paper-and-pencil type—essay questions, short-answer questions, true-false questions, multiple-choice questions

- manipulating apparatus in one way or another, such as using a globe to locate a river

- role-playing a historical situation or portraying the student interpretation of an event

- constructing items such as a model of a railway station or small town to achieve an objective

- presentation through all forms of media — the printed page, films, film strips, tape recorders, oral instructions, physical gestures, or any combination of these

The use of many types of test situations can improve the evaluation process and expand the learning experiences or activities social studies teachers can use to evoke observable reactions in students. Asking a student to paint a picture, to solve a complex problem, to argue the merits of a position, to provide data to support an opinion, and/or to plan and to execute an experiment all constitute test situations.

ROLE OF TESTING IN THE INSTRUCTIONAL PROGRAM

All tests, but particularly teacher-made classroom tests, have an obvious relation to the purposes and content of instruction. They also bear a subtle and covert relation to the students' perceptions of the purposes and content of instruction. To the extent that teachers' actions and decisions are based on tests, they may tell more about the real purposes and content of instruction than anything else.

A classroom test can be more than a means of measurement and evaluation. It can motivate and it can teach. By challenging the student's intellect or by stimulating his or her imagination, a test can make a subject more meaningful or interesting. By requiring that the student think about things in a new way or use previously acquired knowledge to solve unfamiliar problems, a test can give students an appreciation for and understanding of the subject. Because the test is one of the criteria for assessing a student's achievement, it represents to the student direct, material evidence of what has been defined as achievement.

Because classroom tests have an impact on instruction and learning and are likely to measure only what they are constructed to measure, sufficient time devoted to planning a test by social studies teachers is essential. Some helpful questions that teachers should ask themselves before constructing a test are:

- Why is the test being given?
- What will be learned from it?
- What will be learned about each student, the class, the course, or the curriculum?
- How is the test a part of the total evaluation program for this course?

- How does the test relate to the goals of instruction?
- Which objectives will be measured by this test? Which ones excluded?
- How much weight will be given in the test to each of the objectives which are to be measured?
- How will these weights correspond to the relative importance of the objectives in this course?
- Which questions will be concerned with which objectives?
- What level of performance will be expected of the students?
- Is this test a mastery test?
- What kind of response to each question will be expected of the typical student?
- How are these responses related to what the students have been expected to learn?

The teacher's answers to these questions will determine, in a large measure, the character of the test to be developed. They also have an important bearing on the validity and reliability of the test results — particularly on the validity. Since the main purpose of a test is to obtain information, it is at the planning stage of test construction that many of the crucial decisions are made that will determine the nature and the worth of the information that is eventually obtained.

A well-planned test can:

- serve as a vehicle for assisting in determining students' grades
- determine the effectiveness of student learning
- provide the teacher and students with information relating to their attitudes and feelings toward some aspect of social studies content or experience
- provide students with situations which will enable them to "practice" a particular intellectual behavior related to some aspect of social studies

- assess which aspects of the social studies curriculum are useful for the student
- obtain data as to the quality of the learning environment for producing particular types of learning
- assess whether or not students learned a particular list or grouping of information, a process, an attitude, a value, or a skill
- provide diagnostic information as to the level of understanding the students possess prior to investigating a unit or to entering a class in social studies
- provide information for planning specific directions of study in a unit or topic

DETERMINING THE TEST CONTENT

The first step in the construction of any test should be the review of the instructional objectives of the course or unit that students are to achieve. This step can help to control two common faults in any type of test. First, it will eliminate or sharply reduce the number of unrelated or trivia items. Second, it will to some degree, control the "balance" of the test so that certain objectives are not heavily over-represented to the neglect of others. The importance of carefully delineating the subject-matter and making sure the test concentrates on important items which require the student to demonstrate essential skills, understanding, and knowledge cannot be over emphasized. Generally, if some procedure of this sort is not used, test item development will be haphazard, and test results will provide inappropriate "feedback" to the student and teacher for planning further instruction.

A test should not be thought of merely in terms of its content but also in terms of the kinds of responses the test items are capable of evoking from students. Many teachers prefer to construct a content outline or table of specifications to help achieve a balance in content.

A table of specifications, such as the one on p. 16, can be used to determine whether the test will measure a representative sample of the instructional objectives and the content included in the instruction. The instructional objectives are typically listed across the

top of the table and the content areas are listed down the left side. Each objective and each area of content is then reviewed by the teacher in terms of its relative importance to the subject matter and amount of instructional time. In the following table of specifications for a hypothetical test in United States history, numbers are placed in the blocks to indicate the number of test items to be allotted to each objective and each area of content.

TABLE OF SPECIFICATIONS

INSTRUCTIONAL OBJECTIVES / CONTENT AREAS	KNOWLEDGE	TRANSLATION	INTERPRETATION	APPLICATION	ANALYSIS	SYNTHESIS	EVALUATION	TOTAL ITEMS
PRE-REVOLUTIONARY PERIOD	3		1					4
THE REVOLUTION	4	1			5		1	11
THE CONSTITUTION			2	3		1	1	7
WESTWARD EXPANSION	1	1		2	1	2		7
THE CIVIL WAR	3	1			2	1	1	8
RECONSTRUCTION	4		1		1		1	7
TOTAL ITEMS	15	3	4	5	9	4	4	44

As a rule of thumb, it would be desirable to have at least 10 or more items for each specific instructional objective and for each instructional objective content area. However, 10 items in each are not necessary if (1) the area to be tested has been limited and/or (2) the learning outcomes to be measured have been clearly defined. The larger the number of items used, the more dependable the results and the more useful the data become for diagnostic purposes.

Another example of a content outline is the following chart of a hypothetical two-hour, 150-item test in American history. The teacher

decided that 40 percent of the test should be concerned with the recall of factual information, 30 percent with important skills, and 30 percent with the use of critical-thinking abilities. The number of items were determined on the horizontal bar to test the student behaviors while the percentage and number of items on the vertical bar were determined to test the content (knowledge) achievement of the students.

TABLE OF SPECIFICATIONS

CONTENT	BEHAVIOR			TOTAL	
	Factual Information	Skills	Critical Thinking Abilities	Percentage	Number of Items
Colonial Period	10	7	7	16	24
Westward Expansion Period	12	8	7	18	27
Civil War	15	10	11	24	36
Late Nineteenth Century and The Rise of Industrialism	10	9	8	18	27
Twentieth Century	13	11	12	24	36
TOTAL					
Percentage	40	30	30	100	
No. of Items	60	45	45		150

Once a teacher has reviewed the test in terms of the content skill outline, then the teacher must decide what type of test to use that will result in the improvement of instruction and the evaluation process.

TYPES OF TEST

Three kinds of tests are appropriate for evaluating student performance in the social studies: (1) the norm-referenced achievement test, (2) the criterion-referenced test, and (3) the informal teacher-constructed or "home-made" test.

Norm-referenced tests have been standardized by administration to thousands of students, whose scores become the "norms." By referring to a table of appropriate norms furnished by the test publisher, the teacher can convert a student's raw score to a figure that shows how the student stands in *relation to the performance of others.* The conversion may be in terms of percentile ranks (if a student has a percentile score of 74, his or her performance is better than 74 percent of those upon whom the test was standardized); or it may be stated in terms of grade-equivalent scores (if a student has a grade-equivalent score of 4.2, the student performance is comparable to that of the student who is in the second month of Grade 4.)

Norm-referenced tests are designed:

- with a generalized approach to content
- for measuring group achievements and not individual mastery
- to show student relationship to a group and not a mastery of a goal of instruction

Norm-referenced tests are helpful for learning whether or not a student made greater progress last year than this year; whether a student is progressing at all; and whether, in relation to other children of his or her age, the student is well informed and well equipped, or poorly informed and poorly equipped.

Criterion-referenced tests (sometimes called mastery or prescriptive tests) are designed to evaluate a student's progress toward specific objectives. They measure degree of mastery, which may be computed as the percentage of correct items. The incorrect items may indicate what additional instruction is needed to bring about complete mastery.

Criterion-referenced tests are designed to:

- measure an individual student's progress toward specific goals
- indicate the effectiveness of a given program on the individual learner
- monitor the progress of an instructional program toward achieving its objectives by viewing the mastery of the objectives by individual students

The major differences between norm-referenced and criterion-referenced tests are that the former focus on the scores achieved by a large number of students while the latter focus on the individual student's degree of mastery of specific segments of the subject or on the individual student's ability to perform specific skills.

Informal teacher-constructed tests are those tests commonly used by social studies teachers. Whether they involve simple recall of information by requiring students to complete sentences or fill in blanks or are recognition tests of the true-false, multiple choice, or matching varieties, there is no reason why teacher-constructed tests cannot be criterion-referenced tests. To conform to the critical features of criterion testing, each question or test item should relate in a significant way to the specific segments of the subject or the skills that constitute the objectives of the unit. The use of a content outline or a table of specifications will avoid cluttering the test with questions about trivial or inconsequential information.

Social studies educators can derive more information from their tests if they consciously create informal tests which will measure not only factual knowledge but also will assess the mastery of the instructional objectives, strategies, and the materials employed by the teacher to achieve the objectives of the unit or course.

IMPROVEMENT OF SELECTED EVALUATION TOOLS

DEVELOPING BETTER OBJECTIVE TESTS

All teachers evaluate the efforts and progress of their students as a regular part of their work. But many are dissatisfied with the quality and usefulness of the appraisals they make. Teachers are continually reviewing the role that instructional objectives play in improving the written test with an awareness of the problems involved in obtaining a full evaluation of the student. There are several characteristics that many educators feel a test should possess. They are:

- clarity—trivia and ambiguity must be avoided
- consistency—test items and responses should be worded so that the form is the same throughout the test
- accuracy—the desired answers should be the correct ones, and the distractors (wrong answers) should indeed distract

Some other important features that a test should possess are:

- emphasis on skills and concepts—the test items should be written to sample adequately each specific learning outcome in these areas
- emphasis on content and methodology—the test items should be written to measure student mastery of content and instructional strategies covered in the course

Social studies teachers should remember that tests are one of the surest means that students have in uncovering the hidden agenda of the course — discovering what teachers regard as important. Therefore, the type of questions and test items teachers select for inclusion in a test becomes very important in the evaluation process.

Constructing Short-Answer Items

The short-answer form is characterized by the presence of a blank on which the student writes the kind of answer called for by the directions. Three varieties of the short-answer form are:

- The question variety.
  Who invented the cotton gin? (___Eli Whitney___)

- The completion variety.
  The explorers of the Louisiana Territory were (___Lewis and Clark___).

- The identification or association variety.
  After each major city write the state in which it is located.
  New Orleans (___Louisiana___)
  San Antonio (___Texas___)
  Atlanta (___Georgia___)

Short-answer items are primarily useful for measuring knowledge of factual information and are not well adapted to measuring understanding and application. This results in the extraction of a statement from primary or secondary sources which contains a fact, or perhaps a generalization, which the teacher considers significant. This statement, with a key word or words omitted, becomes the short-answer item offered to the student. Such an item is likely to measure whether or not the student remembers the statement as it appears in the text and *not* whether the student comprehends the information.

Guidelines for Writing Short-Answer Items

- State the item so that the answer is limited to a word or brief phrase

- Explicitly state and qualify the item so that only one response is correct, for example:
  When did the war between the States end?(___1865___)

  Other possible answers to this question are: *in the 1860s; when Lee surrendered to Grant,* etc. — all these answers are correct, but the teacher expected the answer 1865. A better way of writing the item is: (___1865___) is the year that Lee surrendered to Grant ending the war between the States.

- Use terms which will have the same meaning to all students

- Make the sentence structure as simple as possible so that the question will be clear to all students

- Ask for important information only
- Make answer blanks equal in length to prevent length as a clue
- Avoid items which have more than one correct answer
- Don't use questions verbatim from the textbook or lecture notes

Constructing Multiple-Choice Items

The multiple-choice item consists of a *stem* that is stated in question or incomplete statement form and several possible answers. Correct response or responses will be called the answer(s), and the incorrect alternatives will be called *distractors*. Their function is to distract those students who have not achieved the specific learning outcome being measured by the item.

Some varieties of the multiple-choice form are:

- The correct-answer variety.
  Who invented the sewing machine? (__b__)
  (a) Singer (c) Whitney
  (b) Howe (d) Fulton

- The best-answer variety.
  What was the basic purpose of the Marshall Plan? (__b__)
  (a) Military defense of Western Europe
  (b) Re-establishment of business and industry in Western Europe
  (c) Settlement of differences with Russia
  (d) Direct help to the hungry and homeless in Europe

  (NOTE: It is important to remember that the difference between the best-answer and correct-answer items is more a difference of topic than of form. The inventor of the sewing machine is recorded in history beyond question or doubt. The purposes of the Marshall Plan cannot be stated with such precision.)

- The multiple-response variety.
  Which of the following are examples of an ethnic group?
  (__b,d__)

  (a) Southerners
  (b) Italian Americans
  (c) Texans
  (d) Anglo Americans

- The incomplete-statement variety.
  The predominate religion in colonial New England was
  (__d__)

  (a) Anglican
  (b) Lutheran
  (c) Quaker
  (d) Congregational
  (e) Baptist

- The negative variety.
  Dictatorships do *not* accept the principle that (a) might makes right, (b) strict censorship is necessary, (c) the state exists for the individual, (d) one political party should be supreme. (__c__)

The multiple-choice item is useful for measuring learner outcomes at the knowledge, understanding, and application levels. Because of its versatility, it is the most widely used of the objective-type items and may follow maps, graphs, verses of poetry, or selections summarizing both sides of some social studies issue. For more information regarding the classification of objective items see "The Objective Test Item" by Harry D. Berg, 35th Yearbook of the National Council of the Social Studies, *Evaluation in Social Studies*.

Guidelines for Writing Multiple-Choice Items

Do

- Write a stem that presents a single, definite problem
- Keep the reading level low (avoid wordiness and complex sentence structure)
- Write concise, unambiguous, and grammatically correct items
- Use questions which are germane to the area being measured

- Keep the alternatives brief and similar in form and grammatically consistent with the stem
- Avoid clues to the correct answer (verbal association, textbook language, length of alternatives, grammatical clues)
- Provide about the same number of keyed responses in each question
- Avoid the alternatives "none of the above" and "all of the above," if possible
- Review each completed item for clarity and for relevance to the specific learning outcome to be measured

DON'T

- Have conflicting frames of reference embedded in the same item
- Have grammatical errors
- Be ambiguous
- Provide superfluous information for the item
- Have a pattern in the rotation of keyed responses (for example, every fourth answer key is E)
- Use unnecessary, technical terminology

Constructing True-False Items

The true-false item may hold the questionable distinction of being the most popular type of test item with classroom teachers and the most criticized regarding its construction and the significance of what it apparently measures. When constructing a true-false item, the teacher should be cognizant of its weaknesses as well as its strengths.

Some of the major weaknesses of the true-false test are:

- Since students have a 50/50 chance of answering correctly by guessing, it is difficult to determine which specific learning outcome has been satisfactorily mastered.

- The item does not provide *diagnostic* information. The false item may be correctly marked on the basis of misinformation by the student. For example:

  (__F__) Abraham Lincoln was the first President of the United States.

  Many students may have remembered that Abraham Lincoln was the first *Republican* President and mark the item true, although the teacher was concerned that the students remember that George Washington was the first President of the United States. Also many students might remember that Abraham Lincoln was the 16th President of the United States. Thus, if a student offers a correction to the item *not anticipated* by the teacher, it may be scored as incorrect even though it may be a defensible answer by the student.

The strengths of the true-false test are:

- It may be rapidly and accurately scored
- The scoring is completely objective — extraneous factors such as the student's dress, neatness, politeness, and writing style have no influence on the teacher's marking
- It can be administered relatively quickly since less time per item is required in marking
- It can be relatively quickly constructed and refined
- The structure is less complex than the structures of other items
- It is best used for students' review and for discussion purposes

The teacher, before developing true-false items, should carefully consider the limitation of this item format and select from among the instructional objectives those that are amenable to measurement with items of this type. The teacher should also consider that the true-false item is adaptable for use in situations where the measurement of the acquisition of factual, noninterpretative information is desired — for example, vocabulary, technical terms, dates, and proper names. A comprehensive sampling of content corresponding to such instructional objectives may be obtained with the true-false test.

Guidelines for Writing True-False Items

Do

- Write statements that can be unquestionably judged as true or false (avoid partly true statements)
- Keep the reading level low (avoid vague terms and complex sentence structures)
- Write concise, unambiguous, and grammatically correct statements
- Use statements which are germane to the area being tested
- Avoid negative statements
- Avoid use of specific determiners (e.g., always, never, sometimes, maybe, etc.)
- Keep the true and false statements approximately equal in length and number (to reduce correct guesses)

DON'T

- Use unnecessary words and phrases
- Have more than one theme in the item
- Have a pattern in the order of the responses
- Use statements directly from the textbook or class notes

Constructing Matching Items

The matching item is a specialized form of the multiple-choice item type in which a set of related stems (premises) shows the same set of alternatives.

An example of this type of item is:

Directions: On the line to the left of each statement in Column I write the letter from Column II with which the item is correctly associated.

Column I	Column II
1. (C) This victory was a major factor in obtaining the alliance between France and the 13 colonies.	A. Bunker Hill B. Yorktown C. Saratoga D. Monmouth E. Trenton
2. (E) After Washington's retreat from New York, this victory renewed hope in the American cause.	
3. (B) For this victory, Washington's troops surrounded the enemy by land and French ships cut off help from the sea.	

The strengths of matching items are:

- They may be used to measure the lower levels of the cognitive domain. Vocabulary, dates, events, and simple relationships can be efficiently and effectively measured.
- They are particularly adaptable to who, what, when, or where situations.
- Higher levels of the cognitive domain can be developed when qualified material is presented and the matching exercise is based on this material.
- They can be scored rapidly, accurately, and objectively — even by individuals who are unqualified to teach in the subject area being examined.

The weaknesses of the matching type of items are:

- They are not particularly applicable to measurement at the higher levels of the cognitive domain.
- The items of a set are interdependent. How the student responds to one premise of a set influences the responses to all other premises.
- It is difficult to have homogeneous premises, the first prerequisite for a set of matching items to function properly.

For example:

The following items refer to the Pre-Revolutionary period in United States history (1763-1775). Match the infor-

mation given under Column I with the appropriate information given under Column II. An answer may be used only one time.

Column I	Column II
1. (B) Regulated colonial trade	A. Parliament
2. (C) Frontier closed by proclamation	B. Navigation Acts
3. (A) English legislative body	C. 1763
4. (F) British troops raided Lexington and Concord	D. Stamp Act
5. (D) Taxed newspapers, legal papers, calendars	E. 1768
	F. 1775

A weakness in this example is the attempt to use premises which are not homogeneous. In order for the items to be used in this manner, the introductory statement was ambiguous. The student knows that the items which follow are related to the period 1763 to 1775. The use of heterogeneous items gives a number of clues to the student. For example, without any knowledge of British government, item 3 should be correctly answered. Alternates "A," "C," "E," and "F" would likely be eliminated from consideration for items 1, 2, 5, and item 4 is likely to be answered by one of three alternates (C,E,F).

Social studies teachers should keep instructional objectives in mind when preparing matching exercises. If this is done, and if the exercise is developed with the intent of choosing the type of item appropriate to the purpose of the test (mastery), the matching exercise should be valid and its items should contribute to the achievement of adequate test reliability.

Guidelines for Writing Matching Items

Do

- Establish a frame of reference for answering the test items
- Establish a general orientation in the introductory statement
- Write clear directions indicating the basis for matching and indicating whether the responses may be used more than once

- Be clear, concise, and use correct grammar
- Keep the material in the list the same (homogeneous)
- Use for who, what, when, or where situations

DON'T

- Provide under directions superfluous information that might confuse the student
- Use unnecessary, technical terminology
- Use items in either column that are not of the same content, as in the following example:

Directions: Select from Column II the name or date which matches an item in Column I. Each answer may be used only one time.

	Column I	Column II
1. (D)	The year in which the Declaration of Independence was signed.	A. George Washington
2. (A)	The first President of the United States	B. Benjamin Franklin
3. (E)	The year in which the Civil War began	C. Hank Aaron
4. (C)	The baseball player who holds the homerun record	D. 1776
5. (B)	The inventor of bifocals	E. 1861

The weakness in this test is found in Column I. Since dates and names are mixed in Column II, choices 1 and 3 have only two alternates available for consideration (D and E). Also, for choices 2, 4, and 5 only three choices are available for consideration. The items in Column II are not the same in content, covering nearly two centuries in time and several unrelated topics (sports, politics, and inventions). Because Columns I and II have the same number of choices, the student is able to answer this test without the teacher ascertaining whether he or she knows the information or not. The test rewards the student's persistence and searching ability more than it does his or her knowledge of history.

Deciding on Types of Test Items

Some guidelines that teachers should consider in deciding which type of test item to use are:

- Use the *short answer type* item when the desired student behavior is to "define," "name," or "describe" something
- Use the *multiple-choice item* when the desired student behavior is to "identify" something, "distinguish" between, or "match" things that are related in some way.
- Use the *true-false item* when the desired student behavior is to distinguish between fact and opinion and between correct and incorrect procedures (where there are only two possible alternatives).
- Use the *matching item* when the desired student behavior is to relate the same things (e.g., dates and events, authors and books, instruments and uses).

These guidelines emphasize the importance of selecting test items which will directly measure each learning outcome and instructional objective. This will improve the informal (home-made) test and will work toward achieving teacher-made criterion-referenced tests to improve and enhance classroom instruction.

DEVELOPING BETTER ESSAY TESTS

Essay type questions require the student to read a question or problem, formulate, and then write the response. This type of question can be used to measure many processes. It can require the student to:

- make comparisons
- supply definitions
- make interpretations
- make evaluations
- explain relationships

The essay examination can be used in assessing the student's higher-order mental processes: application, analysis, synthesis, and evaluation. The essay item can also measure the student's ability to select and organize ideas in writing.

Limitations of the essay type question is that the format:

- Allows the student too much freedom in answering questions
- Hinders the development of questions covering all the content and objectives deemed important by the teacher (content validity)
- Creates a lack of scoring economy as compared to an objective-type test which can be scored in a short period of time
- Lends itself to subjective grading or measurement

The essay question, like the objective test item, must follow certain guidelines to be an effective evaluation tool. They are:

- Major attention should be given to the reasons for the test
- The background and ability of the group to be tested must be considered
- The instructional objectives to be measured must be identified
- The content of the question should be related to an important objective of instruction that can be defined in terms of expected student behavior
- The format of the question should vary with the purpose and content of the question
- The desired student response to each essay question should be outlined or summaized by the teacher so as to have a key for acceptable answers
- The student should know the "ground rules" so that each has the same opportunity to plan and organize his or her time

The social studies teacher should remember that the value of the essay test lies in its potential for the student to demonstrate many of the important social studies skills.

THE ROLE OF VALIDITY AND RELIABILITY IN TESTING

Before a teacher uses test information, there should be confidence that the test measured some specific student performance. Furthermore, what the test measured must be relevant to the uses the teacher intends to make of the information it has provided. This is known as the validity and reliability of a test.

Validity is the degree to which the test measured what the teacher wanted the test to measure, or stated another way, the extent to which the test items adequately sampled the objectives and the subject-matter content of the instructional unit or lesson. A classroom test should be both consistent and relevant to the instructional objectives. This combination of characteristics is called validity.

Reliability is the degree of consistency among test scores. A test is consistent when students who obtain high scores on one set of test items also obtain high scores on other sets of presumably equivalent items, and those who obtain low scores on one set of items also obtain low scores on other sets of similar items.

CHECK LIST FOR EVALUATING INFORMAL CLASSROOM TESTS

As a teacher finishes constructing a test there are several things to be done before the test is administered to a group of students. First, the teacher should allow another teacher to review the test to determine if the test items sample the objectives and subject-matter content.

Once the test has been examined by another professional, the next step is to check the test by some criterion or list. The following suggested check list would be useful in evaluating a classroom test.

Check List

The Test Plan	Yes	No
1. Does the plan include the instructional objectives and the content that will be measured?	____	____

2. Does the table of specifications indicate the emphasis to be given to each instructional objective and to each content area? _____ _____
3. Will the test assess the mastery of the instructional objectives, i.e., skills, knowledge, and understandings? _____ _____

The Test Items **Yes** **No**

4. Are the types of test items appropriate for the learning outcomes to be measured? _____ _____
5. Do the test items measure a representative sample of the instructional objectives and subject-matter content? _____ _____
6. Are the test items of appropriate difficulty for the learning outcomes to be measured? _____ _____
7. Are the test items that measure the same instructional objectives grouped together in the test? _____ _____
8. Is the answer to each test item one that experts (or other teachers) would agree upon? _____ _____
9. Do the test items present clear and definite tasks to be performed? _____ _____
10. Are the test items free from technical defects (clues, ambiguity, inappropriate reading level)? _____ _____
11. Are the items of each type consistent with the rules for constructing that type item? _____ _____
12. Are the test items numbered in a consistent style? _____ _____

The Test **Yes** **No**

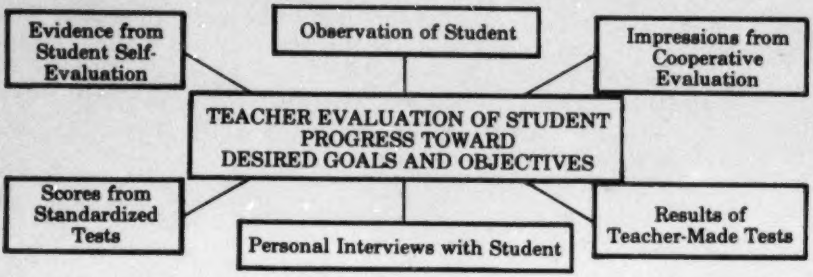
13. Are specific, clear directions provided for each section of the test and for the test as a whole? _____ _____
14. Are the test items arranged in order of increasing difficulty within each section of the test and within the test as a whole to the extent possible? _____ _____
15. Does the test item layout on the page contribute to ease of reading and scoring? _____ _____

USING THE DATA GAINED FROM EVALUATION TO IMPROVE INSTRUCTION

PROVIDING FEEDBACK TO THE INSTRUCTIONAL PROGRAM

Although tests are the most commonly used instruments of evaluation, they are not necessarily the most important or the most useful to the teacher in determining individual or group mastery of instructional goals or objectives deemed important by the teacher. As the following chart shows, a broader concept of evaluation should be the major goal of all teachers.

Evaluation Chart


Evidence from Student Self-Evaluation
Observation of Student
Impressions from Cooperative Evaluation
TEACHER EVALUATION OF STUDENT PROGRESS TOWARD DESIRED GOALS AND OBJECTIVES
Scores from Standardized Tests
Personal Interviews with Student
Results of Teacher-Made Tests

The chart shows an evaluation program that is based on:

- direct student observation
- cooperative evaluation by the teacher and student
- results of teacher-made tests
- personal interviews with the student
- scores from standardized tests
- data from the student self-evaluation

This process gives the social studies teacher viable tools for assessing the student's progress toward desired understandings, attitudes, skills, and behaviors in the social studies.

FUNCTION OF EVALUATION IN THE INSTRUCTIONAL PROGRAM

The responsibility of evaluating student achievement in social studies rests with the teacher. The teacher makes an evaluation of the student and then communicates selected results to the student. This process involves:

- recognizing a need for information
- deciding on the assessment technique
- specifying what is to be evaluated
- collecting and analyzing the data
- comparing results of data with the instructional objectives

The communication part of feedback consists of the following:

- estimating the kinds of evaluation results that the student needs and is ready for and that will most likely facilitate learning
- selecting the evaluation information to present to the student
- deciding on the sequence of presentation
- giving the student the results together with any necessary explanation

The feedback process is often referred to as *cooperative evaluation.* It is a process by which teachers and students critically examine their work together and make judgments about its values and problems. The feedback may involve conveying information about the following to the student:

- the extent to which the student has attained various objectives
- the aptitudes and abilities related to how the student should approach learning tasks

- student behavior patterns or characteristics that affect school learning, such as study habits and perseverance

The feedback may convey information about the following to the teacher:

- student successes in learning particular content
- aspects of the social studies curriculum which are useful for the student
- the quality of the learning environment
- whether students did or did not learn certain information or a process or develop an attitude or a skill
- the levels of understanding students possess prior to investigating a unit or entering a class in social studies
- how to guide the direction of investigation or how to plan the specific directions of study in a unit

Both the teacher and the student can now look at the areas where the student needs help and arrange experiences to help meet the student's needs. Because new social studies experiences grow from and are built on previous experiences, teacher evaluation gives direction to plans for future learning activities. Teacher evaluation also reveals objective evidences that can be shared with parents. Because evaluation is broader in scope than just pencil-and-paper tests, the scope of reporting the evaluation of the students expands. Some examples of information that might be contained in a teacher grade book or a student's folder for evaluation purposes are:

- review of students' notes on films, field trips, etc.
- selected workbook exercises
- selected classroom exercises such as graphs, time lines, and drawings or paintings that summarize social studies experiences or illustrate ideas
- logs of social studies reading, committee work, and special work assignments
- special maps

- written essays
- test papers

In scheduled conferences teacher and parents should study test results, anecdotal records, check lists, and samples of a student's work to discover evidence of interest, initiative, creativity, problems, and needs. The teacher can discuss with the parent the extent and quality of the student's knowledge of useful social studies information. The parent is given an analysis of the student's ability to use various social studies skills. Particular problems that the student and the teacher have encountered are discussed as the teacher refers to notes taken during the period of learning. When the teacher can discuss with parents (using firsthand evidence) the student's performance and achievement in an area of learning, parents can be more objective in evaluating their child's growth and can acquire a keener insight into the objectives of the social studies program.

Thus, the importance of evaluation and feedback in social studies must not be minimized. It goes beyond teacher-student evaluation of stated objectives and shared feedback. It extends to other important aspects of social studies—to progress toward the student's development of new understandings, attitudes, behaviors, and skills; to selection of further experiences to meet identified needs; to accumulation of evaluative data to be shared with parents; and to development of the social studies curriculum on a long-range basis.

APPENDIX A

Numerous articles, books, and parts of books deal with evaluation in the social studies. The following references are not intended to be comprehensive but to identify some publications which may be useful for teachers interested in improving their knowledge about evaluation. The list of General References would be of value in further study about the role of evaluation in the social studies. The list of Selected References would be of importance to teachers interested in improving their evaluation tools.

GENERAL REFERENCES

Buros, Oscar K., ed. MENTAL MEASUREMENTS YEARBOOK, 5th, 6th, 7th eds. Highland Park, NJ: Gryphon Press, 1959, 1965, 1972.

Dizney, Henry. CLASSROOM EVALUATION FOR TEACHERS. Dubuque, IA: William C. Brown Co., 1971.

Kenworthy, Leonard S. SOCIAL STUDIES FOR THE SEVENTIES. Waltham, MA; 2nd ed. New York: Wiley, 1973.

Kurfman, Dana G., ed. EVALUATION IN GEOGRAPHIC EDUCATION. 1971 Yearbook of the National Council for Geographic Education. Belmont, CA: Fearon Publishers, 1971.

Texas Education Agency. Division of Curriculum Development. CONCEPT TEACHING IN SOCIAL STUDIES. Bulletin 759. Austin: The Texas Education Agency, 1975.

Texas Education Agency. Division of Curriculum Development. CLASSROOM STRATEGIES FOR INDIVIDUALIZING INSTRUCTION IN THE SOCIAL STUDIES. Bulletin 770. Austin: The Texas Education Agency, 1976.

SELECTED REFERENCES

Berg, Harry D., ed. EVALUATION IN SOCIAL STUDIES. 35th Yearbook. Washington, D.C.: National Council for the Social Studies, 1965.

Bloom, B. S. and others. TAXONOMY OF EDUCATIONAL OBJECTIVES, HANDBOOK I: COGNITIVE DOMAIN. New York: David McKay Co., 1956.

Brown, Lena B., ed. "Testing in the Social Studies: Practical Ideas for Classroom Teachers." SOCIAL EDUCATION. Journal of the National Council for the Social Studies, Vol. 40, No. 7 (November-December 1976), pp. 503-541.

Gronlund, Norman E. PREPARING CRITERION-REFERENCED TESTS FOR CLASSROOM INSTRUCTION. New York: Macmillan, 1973.

Hunkins, Francis P. QUESTIONING STRATEGIES AND TECHNIQUES. Boston: Allyn and Bacon, Inc., 1972. 140 pp.

Krathwohl, D. R.; Bloom, B. S.; and Masia, B. B. TAXONOMY OF EDUCATIONAL OBJECTIVES: AFFECTIVE DOMAIN. New York: David McKay Co., 1964.

Mager, Robert F. MEASURING INSTRUCTIONAL INTENT. Belmont, CA: Fearon Publishers, 1973.

Marshall, John C. and Hales, Loyde W. CLASSROOM TEST CONSTRUCTION. Reading, MA: Addison-Wesley Publishing Co., 1971.

Popham, W. James. EVALUATING INSTRUCTION. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1973.

Sanders, Norris M. CLASSROOM QUESTIONS: WHAT KIND? New York: Harper and Row, 1966.

APPENDIX B

GLOSSARY OF TERMS

Affective Domain—a broad classification of educational outcomes involving responses or mental tendencies such as acceptance, rejection, the giving of attention, interests, attitudes, appreciations, values, emotional sets, biases, and qualities of character or conscience

Assessment—a variety of measurements and procedures that seek to describe how well students are doing relative to a predetermined set of goals in various fields. It is a term frequently associated with local, state, and national projects.

Need Assessment—a method of determining if a discrepancy exists between a predetermined quality of performance and the actual performance, and establishing the degree of discrepancy

Cognitive Domain—a broad classification of educational outcomes with primary emphasis on the mental or intellectual processes such as knowledge, understanding, intellectual abilities, and intellectual skills

Cooperative Evaluation—process by which teacher and students critically examine their work together and make some judgments about its values and problems

Criterion—rule or standard used as a test of quality

Criterion-Referenced—a method of determining accomplishment by comparing a student's demonstrated performance of a specific skill or grasp of a segment of knowledge with an established standard of quality (criterion)

Curriculum—the strategies and all the planned sequences of learning activities for attaining the objectives of the educational program. It includes not only the course of study but also the goals and even the extracurricular activities if they are designed to contribute to education purposes.

Diagnostic Testing—a means of documenting student needs through measurement instruments

Evaluation—the process of examining and judging the fulfillment of an activity or series of activities intended to produce a product (regardless of the magnitude of the intended product). In education, at any level, evaluation may be of students or of instructional programs.

Grading—a system for providing a tangible estimate of an individual's accomplishments (generally associated with a given amount of expected effort within an established time frame, although these factors are not necessarily pertinent). Grading may take the form of letters or numerals.

Meaningful Objective—an objective stated in such a way that each person who reads it can understand clearly what is to be accomplished to complete it

Measurement—the reflected result by which student behavior or characteristics may be evaluated

Measurable Objective—an objective stated in such a way that each person who reads it can visualize a practical and effective way to measure the extent to which the objective was accomplished

Norm—a standard or average level of performance for a specific group on a particular task or tests. For example, the median score of a nationwide sample is often used as the norm for a norm-referenced test. It should be noted that such a norm is a typical level of performance, not necessarily the ideal level.

Norm-Referenced Test—an instrument determining accomplishment by comparison of a student test score with a set standard of average achievement previously established from the scores of a wide spectrum of students

Objective—a desired outcome of educational experiences

Objective Test—a term that has come to be applied to those tests in which a person marks a correct answer or fills in a blank with a correct answer. The correct answer to each question has been determined in advance by a consensus of judges. The label "objective" identifies a particular kind of test format; it does not mean that the test is free of bias, error, or arbitrary decisions of correctness. (Strictly speaking, there is no such thing as an objective test in relation to objectivity.)

Reliability—the degree to which a test consistently reflects the same results. If a test is reliable, it will yield very nearly the same score for a student if it is administered to the student several times within a short period of time.

Standardized Test—a test in which all students answer a number of questions under the same directions and time conditions

Stem, Key, and Distractors—the actual test problem is called the **stem**, and the correct answer or choice is called the **key**. If there are wrong answers given, as in a multiple-choice item, they are called **distractors**.

Summative Evaluation—evaluation that combines the product evaluation (the extent to which objectives have been met) with information gathered during the formative evaluation to arrive at a final group of statements assessing the success of product

Taxonomy—a way of organizing and naming things (a classification system)

Testing—the process of evaluating student achievement

Testing Items—the problems or questions constructed for the student. The common forms of objective test items are true-false, multiple-choice, matching, and fill-in or completion.

Validity—the degree to which a test measures what it is supposed to measure. Validity can be defined in terms of content. (Does a test deal with what is taught and in a manner similar to the way in which it is taught?) Another way of determining validity is to establish some criterion to measure educational objectives and then to determine the relation of the test to that criterion.